DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A MODEL FOR BAYESIAN SOURCE SEPARATION WITH THE OVERALL MEAN

Daniel B. Rowe

SOCIAL SCIENCE WORKING PAPER 1118

April 2001

A Model for Bayesian Source Separation With The Overall Mean

Daniel B. Rowe
H & SS 228-77
California Institute of Technology
Pasadena, California 91125

Abstract

Typically in source separation models the overall mean as well as the mean of the sources are assumed to be zero. This paper assumes a nonzero overall mean and a nonzero source mean, quantifies available prior knowledge regarding them and the other parameters. This prior knowledge is incorporated into the inferences along with the current data in the Bayesian approach to source separation. Vague, conjugate normal, and generalized conjugate normal distributions are used to quantify knowledge for the overall mean vector. Algorithms for estimating the parameters of the model from the joint posterior distribution are derived an determined statistically from the posterior distribution using both Gibbs sampling a Markov chain Monte Carlo method and the iterated conditional modes algorithm a deterministic optimization technique for marginal mean and maximum a posterior estimates respectively. This is a methodological paper which outlines the model without the use of a numerical example.

1 Introduction and Model

The source separation problem is that of separating unobservable or latent source signals when mixed signals are observed. To take a set of observed mixed signal vectors and unmix or separate them into a set of true unobservable source signal vectors. This paper adopts a multivariate Bayesian [8,9] statistical approach and the linear synthesis model [6,10,11,12] with an overall mean.

For motivation and illustation of the source separation model, the context of the "cocktail party problem" is adopted [3]. At a cocktail party, there are p microphones that record or observe m partygoers or speakers at n time increments. The observed conversations consist of mixtures of true conversations. In other words, p-dimensional mixed signal vectors $x_i = (x_{i1}, \ldots, x_{ip})'$ are observed and the goal is to separate these observed signal vectors into m-dimensional true underlying source signal vectors, $s_i = (s_{i1}, \ldots, s_{im})'$ where $i = 1, \ldots, n$.

The linear synthesis source separation model for the observed vector x_i at time i is

$$\begin{array}{ccccccccc} (x_i|\mu,\Lambda,s_i) & = & \mu & + & \Lambda & s_i & + & \epsilon_i, \\ (p\times 1) & & (p\times 1) & & (p\times m) & (m\times 1) & & (p\times 1) \end{array} \tag{1.1}$$

where it has been assumed that the observed signals have a nonzero mean which is retained and determined statistically. The variables in the model are denoted as follows,
μ = the p-dimensional overall mean, $\mu = (\mu_1, \ldots, \mu_p)'$;
Λ = a $p \times m$ matrix of unobserved mixing constants, $\Lambda = (\lambda_1', \ldots, \lambda_p')'$;
s_i = the i^{th} m-dimensional unobservable source vector, $s_i = (s_{i1}, \ldots, s_{1m})'$; and
ϵ_i = the p-dimensional vector of errors or noise terms of the i^{th} observed signal vector $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})'$.

2 Likelihood

It is specified that the errors of the observations are independent over time and motivated by the central limit theorem, normally distributed random vectors with mean zero and covariance matrix Ψ. Thus, the likelihood of a given observation vector is

$$p(x_i|\mu,\Lambda,s_i,\Psi) \propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(x_i-\mu-\Lambda s_i)'\Psi^{-1}(x_i-\mu-\Lambda s_i)}. \tag{2.1}$$

The joint likelihood of the observations is

$$p(x_1,\ldots,x_n|\mu,\Lambda,s_1,\ldots,s_n,\Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu-\Lambda s_i)'\Psi^{-1}(x_i-\Lambda s_i)}. \tag{2.2}$$

Analogous to regression, the source separation model can be written in terms of matrices as

$$\begin{array}{ccccccccc} (X|\mu,\Lambda,S) & = & e_n\mu' & + & S & \Lambda' & + & E & , \\ (n\times p) & & (n\times p) & & (n\times m) & (m\times p) & & (n\times p) & \end{array} \tag{2.3}$$

with likelihood

$$p(X|\mu,\Lambda,S,\Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \tag{2.4}$$

where $X' = (x_1, \ldots, x_n)$ contains the observations as rows of X, e_n is the n dimensional unit vector, $S' = (s_1, \ldots, s_n)$ contains the unobserved true source vectors as rows of S, and $E' = (\epsilon_1, \ldots, \epsilon_n)$contains the error vectors as rows of E. The time series of observations for the j^{th} microphone is the j^{th} column of X and the time series of unobservables for the k^{th} source is the k^{th} column of S.

Available knowledge regarding how probable values of the parameters are in the form of prior distributions is now quantified and incorporated into the inferences.

3 Priors And Posteriors

The mean is specified to be either vague, conjugate normally distributed with mean μ_0 and covariance matrix $\psi_0\Psi$, or generalized conjugate normally distributed with mean μ_0 and covariance matrix Δ. The source vectors s_i are specified to be normally distributed with mean s_{i0} and covariance matrix R. Regarding the other parameters, information is incorporated as an inverted Wishart distribution for the covariance of the source vectors with η degrees of freedom and scale matrix V, an inverted Wishart distribution for the covariance of the observed vectors with ν degrees of freedom and scale matrix Q. The mixing matrix is specified to be either conjugate normally distributed with mean $\lambda_0 = vec(\Lambda_0')$ and covariance $\Psi \otimes A$ or generalized conjugate normally distributed with mean $\lambda_0 = vec(\Lambda_0')$ and covariance Δ. The inverted Wishart distribution is the multivariate generalization of the inverted gamma distribution which is used as a prior distribution for variances.

The vague, conjugate normal, and generalized conjugate normal prior distributions are

$$p(\mu) \propto \text{a constant} \tag{3.5}$$
$$p(\mu|\Psi) \propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'(\psi_0\Psi)^{-1}(\mu-\mu_0)} \tag{3.6}$$
$$p(\mu) \propto |\Gamma|^{-\frac{1}{2}} e^{(\mu-\mu_0)'\Gamma^{-1}(\mu-\mu_0)} \tag{3.7}$$

respectively. In addition, prior information regarding the mixing matrix is quantified using the conjugate and generalized conjugate prior distributions

$$p(\Lambda|\Psi) \propto |A|^{-\frac{p}{2}}|\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \tag{3.8}$$
$$p(\lambda) \propto |\Delta|^{-\frac{1}{2}} e^{-\frac{1}{2}(\lambda-\lambda_0)'\Delta^{-1}(\lambda-\lambda_0)}. \tag{3.9}$$

Prior distributions are assessed for the remaining model parameters. It is specified that the prior distributions for the sources S, the source covariance matrix R, the error covariance matrix Ψ, follow normal, inverted Wishart, and inverted Wishart distributions respectively

$$p(S|R) \propto |R|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'}, \tag{3.10}$$
$$p(R) \propto |R|^{-\frac{\eta}{2}} e^{-\frac{1}{2}trR^{-1}V}, \tag{3.11}$$
$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}Q}, \tag{3.12}$$

where $\nu > 2p$, $\eta > 2m$. The hyperparameters μ_0, ψ_0, Γ, Λ_0, A, Δ, S_0, η, V, ν, and Q are to be assessed.

Note that both Ψ and R are full covariance matrices allowing the elements of the observed mixed signal and the unobserved source component vectors to be correlated or dependent.

Upon using Bayes' rule the posterior distributions for the unknown parameters, taking either the vague, conjugate normal, or generalized conjugate normal priors for the overall mean are

$$
p(\mu,S,R,\Lambda,\Psi|X)\propto\begin{cases}
|\Psi|^{-\frac{(n+\nu+m)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}U_1} \\
\times|R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]}, & (3.13)\\
\\
|\Psi|^{-\frac{(n+\nu+m+1)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}U_2} \\
\times|R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]}, & (3.14)\\
\\
|\Psi|^{-\frac{(n+\nu)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}[(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')+Q]} \\
\times|R|^{-\frac{(n+\eta)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \\
\times|\Gamma|^{-\frac{1}{2}}e^{(\mu-\mu_0)'\Gamma^{-1}(\mu-\mu_0)}|\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda-\lambda_0)\Delta^{-1}(\lambda-\lambda_0)'}, & (3.15)
\end{cases}
$$

where

$$
\begin{aligned}
U_1 &= (X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')+(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)' \\
&\quad +Q \qquad (3.16)
\end{aligned}
$$

and

$$
\begin{aligned}
U_2 &= (X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')+(\mu-\mu_0)\psi_0^{-1}(\mu-\mu_0)' \\
&\quad +(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'+Q. \qquad (3.17)
\end{aligned}
$$

These posterior distributions are now evaluated in order to obtain parameter estimates of the sources, the overall population mean, the mixing matrix, the source covariance matrix, and the errors covariance matrix.

4 Vague Estimation

With the above posterior distribution, it is not possible to obtain marginal distributions and thus marginal estimates for any of the parameters in an analytic closed form. It is also not possible to find analytic closed form solutions for maximum a posteriori estimates. It is possible to use both Gibbs sampling, a Monte Carlo integration technique to obtain marginal parameter estimates [1,2] and the deterministic optimization technique iterated conditional modes (ICM) for maximum a posteriori estimates [5,7]. For both estimation procedures, the posterior conditional distributions are required.

4.1 Posterior Conditionals

From the joint posterior distribution we can obtain the posterior conditional distributions.

The conditional posterior density of the overall mean is

$$\begin{aligned}
p(\mu|S,R,\Lambda,\Psi,X) &\propto p(\mu)p(X|\mu,S,\Lambda,\Psi) \\
&\propto \text{(a constant)} \\
&\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr(X-e_n\mu'-S\Lambda')\Psi^{-1}(X-e_n\mu'-S\Lambda')'} \\
&\propto e^{-\frac{1}{2}(\mu-\tilde{\mu})'\left(\frac{\Psi}{n}\right)^{-1}(\mu-\tilde{\mu})}
\end{aligned} \tag{4.1}$$

where the posterior conditional mean and mode is given by

$$\tilde{\mu} = \bar{x} - \Lambda\bar{s}. \tag{4.2}$$

That is, the overall mean given the other parameters and the data is normally distributed. Note that the estimator of the overall mean is not the sample mean. The estimator of the overall mean should not be the sample mean especially when the sources do not have a mean of zero or in "small" samples. If the estimator is the sample mean, then it is a biased estimator.

The conditional posterior distributions for the mixing matrix is

$$\begin{aligned}
p(\Lambda|\mu,S,R,\Psi,X) &\propto p(\Lambda|\Psi)p(X|\mu,\Lambda,S,\Psi) \\
&\propto |\Psi|^{-\frac{m}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \\
&\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\
&\propto e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\tilde{\Lambda})(A^{-1}+S'S)(\Lambda-\tilde{\Lambda})'}
\end{aligned} \tag{4.3}$$

where the posterior conditional mean and mode is given by

$$\tilde{\Lambda} = [(X-e_n\mu')'S + \Lambda_0 A^{-1}](A^{-1}+S'S)^{-1}. \tag{4.4}$$

The conditional distribution for the mixing matrix given the other parameters and the data is normally distributed.

The conditional posterior distribution of the observation error matrix is

$$\begin{aligned}
p(\Psi|\mu,S,R,\Lambda,X) &\propto p(\Psi)p(\Lambda|\Psi)p(X|\mu,S,\Lambda,\Psi) \\
&\propto |\Psi|^{-\frac{\nu}{2}}e^{-\frac{1}{2}tr\Psi^{-1}Q}|\Psi|^{-\frac{m}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \\
&\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\
&\propto |\Psi|^{-\frac{(n+\nu+m)}{2}}e^{-\frac{1}{2}tr\Psi^{-1}U_1}
\end{aligned} \tag{4.5}$$

where

$$\begin{aligned}
U_1 &= (X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda') \\
&\quad +(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)' + Q
\end{aligned} \tag{4.6}$$

with a mode given by

$$\tilde{\Psi} = \frac{U_1}{n+\nu+m}\ . \tag{4.7}$$

The conditional distribution of the observation error covariance matrix given the other parameters and the data is an inverted Wishart.

The conditional posterior distribution for the sources is

$$\begin{aligned} p(S|\mu, R, \Lambda, \Psi, X) &\propto p(S|R)p(X|\mu, \Lambda, S, \Psi) \\ &\propto |R|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\ &\quad \times |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\ &\propto e^{-\frac{1}{2}tr(S-\tilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\tilde{S})'} \end{aligned} \tag{4.8}$$

where the posterior conditional mean and mode is given by

$$\tilde{S} = [(X - e_n\mu')\Psi^{-1}\Lambda + S_0 R^{-1}](R^{-1} + \Lambda'\Psi^{-1}\Lambda)^{-1}. \tag{4.9}$$

The conditional posterior distribution for the sources given the other parameters and the data is normally distributed.

The conditional posterior distribution for the source covariance matrix is

$$\begin{aligned} p(R|\mu, \Lambda, S, \Psi, X) &\propto p(R)p(S|R)p(X|\mu, \Lambda, S, \Psi) \\ &\propto |R|^{-\frac{\eta}{2}} e^{-\frac{1}{2}trR^{-1}V} |R|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\ &\propto |R|^{-\frac{(n+\eta)}{2}} e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \end{aligned} \tag{4.10}$$

with the posterior conditional mode given by

$$\tilde{R} = \frac{(S - S_0)'(S - S_0) + V}{n + \eta}. \tag{4.11}$$

The conditional posterior distribution for the source covariance matrix given the other parameters and the data is inverted Wishart distributed.

4.2 Gibbs Sampling

For Gibbs estimation of the posterior, start with initial values for μ, S and Ψ say $\bar{\mu}_{(0)}$, $\bar{S}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned} \bar{\Lambda}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\Lambda|\bar{\mu}_{(l)}, \bar{\mathrm{S}}_{(l)}, \bar{\mathrm{R}}_{(l)}, \bar{\Psi}_{(l)}, \mathrm{X}) \\ \bar{\Psi}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\Psi|\bar{\mu}_{(l)}, \bar{\mathrm{S}}_{(l)}, \bar{\mathrm{R}}_{(l)}, \bar{\Lambda}_{(l+1)}, \mathrm{X}) \\ \bar{R}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mathrm{R}|\bar{\mu}_{(l)}, \bar{\mathrm{S}}_{(l)}, \bar{\Lambda}_{(l+1)}, \bar{\Psi}_{(l+1)}, \mathrm{X}) \\ \bar{S}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mathrm{S}|\bar{\mu}_{(l)}, \bar{\mathrm{R}}_{(l+1)}, \bar{\Lambda}_{(l+1)}, \bar{\Psi}_{(l+1)}, \mathrm{X}) \\ \bar{\mu}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mu|\bar{\mathrm{S}}_{(l+1)}, \bar{\mathrm{R}}_{(l+1)}, \bar{\Lambda}_{(l+1)}, \bar{\Psi}_{(l+1)}, \mathrm{X}) \end{aligned}$$

and the first random variates called the "burn in" are discarded compute from the next L variates

$$\bar{S} = \frac{1}{L}\sum_{l=1}^{L} \bar{S}_{(l)} \qquad \bar{R} = \frac{1}{L}\sum_{l=1}^{L} \bar{R}_{(l)} \qquad \bar{\Lambda} = \frac{1}{L}\sum_{l=1}^{L} \bar{\Lambda}_{(l)}$$

$$\bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Psi}_{(l)} \qquad \bar{\mu} = \frac{1}{L}\sum_{l=1}^{L}\bar{\mu}_{(l)}$$

which are the sampling based marginal posterior mean estimates of the parameters.

4.3 Maximum A Posteriori

The ICM estimaton procedure consists of starting with initial values for μ and S say $\tilde{\mu}_{(0)}$ and $\tilde{S}_{(0)}$ then iterating through

$$\begin{aligned}
\tilde{\Lambda}_{(l+1)} &= [(X - e_n\tilde{\mu}'_{(l)})'\tilde{S}_{(l)} + \Lambda_0 A^{-1}](A^{-1} + \tilde{S}'_{(l)}\tilde{S}_{(l)})^{-1} \\
\tilde{\Psi}_{(l+1)} &= [(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)})'(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)}) + \\
&\quad (\tilde{\Lambda}_{(l+1)} - \Lambda_0)A^{-1}(\tilde{\Lambda}_{(l+1)} - \Lambda_0)' + Q]/(n + \nu + m) \\
\tilde{R}_{(l+1)} &= \frac{(\tilde{S}_{(l)} - S_0)'(\tilde{S}_{(l)} - S_0) + V}{n + \eta} \\
\tilde{S}_{(l+1)} &= (X - e_n\tilde{\mu}'_{(l)})\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)}(\tilde{R}^{-1}_{(l+1)} + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)})^{-1} \\
\tilde{\mu}_{(l+1)} &= \bar{x} - \tilde{\Lambda}_{(l+1)}\bar{\tilde{s}}_{(l+1)}
\end{aligned}$$

until convergence is reached. The converged values $(\tilde{\mu}, \tilde{S}, \tilde{R}, \tilde{\Lambda}, \tilde{\Psi})$ are joint posterior modal (maximum a posteriori) estimators of the parameters.

5 Conjugate Estimation

With the posterior distribution, it is not possible to obtain marginal distributions and thus marginal estimates for any of the parameters in an analytic closed form. It is also not possible to find analytic closed form solutions for maximum a posteriori estimates. It is possible to use both Gibbs sampling, a Monte Carlo integration technique to obtain marginal parameter estimates [1,2] and the deterministic optimization technique iterated conditional modes (ICM) for maximum a posteriori estimates [5,7]. For both estimation procedures, the posterior conditional distributions are required.

5.1 Posterior Conditionals

From the joint posterior distribution we can obtain the posterior conditional distributions.

The conditional posterior density of the overall mean is

$$p(\mu|S, R, \Lambda, \Psi, X) \;\propto\; p(\mu|\Psi)p(X|\mu, S, \Lambda, \Psi)$$

$$
\begin{aligned}
&\propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'(\psi_0\Psi)^{-1}(\mu-\mu_0)} \\
&\quad \times|\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(X-e_n\mu'-S\Lambda')\Psi^{-1}(X-e_n\mu'-S\Lambda')'} \\
&\propto e^{-\frac{1}{2}(\mu-\tilde{\mu})'\left[\frac{n\psi_0}{(1+n\psi_0)}\Psi\right]^{-1}(\mu-\tilde{\mu})}
\end{aligned} \tag{5.1}
$$

where the posterior conditional mean and mode is given by

$$
\tilde{\mu} = \frac{1}{(1+n\psi_0)}\mu_0 + \frac{n\psi_0}{(1+n\psi_0)}(\bar{x} - \Lambda\bar{s}) \tag{5.2}
$$

That is, the overall mean given other parameters and the data is normally distributed.

The conditional posterior distributions for the mixing matrix is

$$
\begin{aligned}
p(\Lambda|\mu, S, R, \Psi, X) &\propto p(\Lambda|\Psi)p(X|\mu, \Lambda, S, \Psi) \\
&\propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \\
&\quad \cdot |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\
&\propto e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\tilde{\Lambda})(A^{-1}+S'S)(\Lambda-\tilde{\Lambda})'}
\end{aligned} \tag{5.3}
$$

where the posterior conditional mean and mode is given by

$$
\tilde{\Lambda} = [(X - e_n\mu')'S + \Lambda_0 A^{-1}](A^{-1} + S'S)^{-1}. \tag{5.4}
$$

The conditional distribution for the mixing matrix given the other parameters and the data is normally distributed.

The conditional posterior distribution of the observation error matrix is

$$
\begin{aligned}
p(\Psi|\mu, S, R, \Lambda, X) &\propto p(\Psi)p(\Lambda|\Psi)p(X|\mu, S, \Lambda, \Psi) \\
&\propto |\Psi|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mu-\mu_0)'(\psi_0\Psi)^{-1}(\mu-\mu_0)} \\
&\quad \times|\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}Q} |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \\
&\quad \times|\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\
&\propto |\Psi|^{-\frac{(n+\nu+m+1)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}G}
\end{aligned} \tag{5.5}
$$

where

$$
\begin{aligned}
G &= (X - e_n\mu' - S\Lambda')'(X - e_n\mu' - S\Lambda') \\
&\quad +(\mu - \mu_0)\psi_0^{-1}(\mu - \mu_0)' + (\Lambda - \Lambda_0)A^{-1}(\Lambda - \Lambda_0)' + Q
\end{aligned} \tag{5.6}
$$

with a mode given by

$$
\tilde{\Psi} = \frac{G}{n + \nu + m + 1} \ . \tag{5.7}
$$

The conditional distribution of the observation error covariance matrix given the other parameters and the data is an inverted Wishart.

The conditional posterior distribution for the sources is

$$
\begin{aligned}
p(S|\mu, R, \Lambda, \Psi, X) &\propto p(S|R)p(X|\mu, \Lambda, S, \Psi) \\
&\propto |R|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\
&\quad \times |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \\
&\propto e^{-\frac{1}{2}tr(S-\tilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\tilde{S})'} \qquad (5.8)
\end{aligned}
$$

where the posterior conditional mean and mode is given by

$$
\tilde{S} = [(X - e_n\mu')\Psi^{-1}\Lambda + S_0R^{-1}](R^{-1} + \Lambda'\Psi^{-1}\Lambda)^{-1}. \qquad (5.9)
$$

The conditional posterior distribution for the sources given the other parameters and the data is normally distributed.

The conditional posterior distribution for the source covariance matrix is

$$
\begin{aligned}
p(R|\mu, \Lambda, S, \Psi, X) &\propto p(R)p(S|R)p(X|\mu, \Lambda, S, \Psi) \\
&\propto |R|^{-\frac{\eta}{2}} e^{-\frac{1}{2}trR^{-1}V} |R|^{-\frac{n}{2}} e^{-\frac{1}{2}tr(S-S_0)R^{-1}(S-S_0)'} \\
&\propto |R|^{-\frac{(n+\eta)}{2}} e^{-\frac{1}{2}trR^{-1}[(S-S_0)'(S-S_0)+V]} \qquad (5.10)
\end{aligned}
$$

with the posterior conditional mode given by

$$
\tilde{R} = \frac{(S - S_0)'(S - S_0) + V}{n + \eta}. \qquad (5.11)
$$

The conditional posterior distribution for the source covariance matrix given the other parameters and the data is inverted Wishart distributed.

5.2 Gibbs Sampling

For Gibbs estimation of the posterior, start with initial values for μ, S and Ψ say $\bar{\mu}_{(0)}$, $\bar{S}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$
\begin{aligned}
\bar{\Lambda}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\Lambda|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l})}, \bar{\Psi}_{(\mathrm{l})}, \mathrm{X}) \\
\bar{\Psi}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\Psi|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l})}, \bar{\Lambda}_{(\mathrm{l+1})}, \mathrm{X}) \\
\bar{R}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mathrm{R}|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\Lambda}_{(\mathrm{l+1})}, \bar{\Psi}_{(\mathrm{l+1})}, \mathrm{X}) \\
\bar{S}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mathrm{S}|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l+1})}, \bar{\Lambda}_{(\mathrm{l+1})}, \bar{\Psi}_{(\mathrm{l+1})}, \mathrm{X}) \\
\bar{\mu}_{(l+1)} &= \text{a random variate from } \mathrm{p}(\mu|\bar{\mathrm{S}}_{(\mathrm{l+1})}, \bar{\mathrm{R}}_{(\mathrm{l+1})}, \bar{\Lambda}_{(\mathrm{l+1})}, \bar{\Psi}_{(\mathrm{l+1})}, \mathrm{X})
\end{aligned}
$$

and the first random variates called the "burn in" are discarded compute from the next L variates

$$
\bar{S} = \frac{1}{L}\sum_{l=1}^{L} \bar{S}_{(l)} \qquad \bar{R} = \frac{1}{L}\sum_{l=1}^{L} \bar{R}_{(l)} \qquad \bar{\Lambda} = \frac{1}{L}\sum_{l=1}^{L} \bar{\Lambda}_{(l)}
$$

$$\bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L}\bar{\Psi}_{(l)} \qquad \bar{\mu} = \frac{1}{L}\sum_{l=1}^{L}\bar{\mu}_{(l)}$$

which are the sampling based marginal posterior mean estimates of the parameters.

5.3 Maximum A Posteriori

The ICM estimaton procedure consists of starting with initial values for μ and S say $\tilde{\mu}_{(0)}$ and $\tilde{S}_{(0)}$ then iterating through

$$\begin{aligned}
\tilde{\Lambda}_{(l+1)} &= [(X - e_n\tilde{\mu}'_{(l)})'\tilde{S}_{(l)} + \Lambda_0 A^{-1}](A^{-1} + \tilde{S}'_{(l)}\tilde{S}_{(l)})^{-1} \\
\tilde{\Psi}_{(l+1)} &= [(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)})'(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)}) + \\
&\quad (\tilde{\Lambda}_{(l+1)} - \Lambda_0)A^{-1}(\tilde{\Lambda}_{(l+1)} - \Lambda_0)' + (\tilde{\mu}_{(l)} - \mu_0)(\psi_0)^{-1}(\tilde{\mu}_{(l)} - \mu_0)' + \\
&\quad Q]/(n + \nu + m + 1) \\
\tilde{R}_{(l+1)} &= \frac{(\tilde{S}_{(l)} - S_0)'(\tilde{S}_{(l)} - S_0) + V}{n + \eta} \\
\tilde{S}_{(l+1)} &= (X - e_n\tilde{\mu}'_{(l)})\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)}(\tilde{R}^{-1}_{(l+1)} + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)})^{-1} \\
\tilde{\mu}_{(l+1)} &= \frac{1}{(1 + n\psi_0)}\mu_0 + \frac{n\psi_0}{(1 + n\psi_0)}(\bar{x} - \tilde{\Lambda}_{(l+1)}\bar{\tilde{s}}_{(l+1)})
\end{aligned}$$

until convergence is reached. The converged values $(\tilde{\mu}, \tilde{S}, \tilde{R}, \tilde{\Lambda}, \tilde{\Psi})$ are joint posterior modal (maximum a posteriori) estimators of the parameters.

6 Generalized Conjugate Estimation

With the generalized natural conjugate normal distribution for the overall mean μ, it is not possible to obtain all or any of the marginal distributions and thus marginal estimates in closed form. It is also not possible to find analytic closed form solutions for maximum a posteriori estimates. It is possible to use both Gibbs sampling, a Monte Carlo integration technique to obtain marginal parameter estimates [1,2] and the deterministic optimization technique iterated conditional modes (ICM) for maximum a posteriori estimates [5,7]. For this reason, marginal estimates are found using the Gibbs sampling algorithm.

6.1 Posterior Conditionals

Both Gibbs sampling and ICM require the posterior conditionals. Gibbs sampling requires the conditionals for the generation of random variates while ICM requires them for maximization by cycling through their modes.

The conditional posterior distribution of the overall mean is

$$\begin{aligned} p(\mu|S,R,\Lambda,\Psi,X) &\propto p(\mu)p(X|\mu,S,\Lambda,\Psi) \\ &\propto |\Gamma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\mu-\mu_0)'\Gamma^{-1}(\mu-\mu_0)} \\ &\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr(X-e_n\mu'-S\Lambda')\Psi^{-1}(X-e_n\mu'-S\Lambda')'} \end{aligned} \tag{6.1}$$

which after some algebra can be written as

$$p(\mu|S,R,\Lambda,\Psi,X) \propto e^{-\frac{1}{2}(\mu-\tilde{\mu})'\left[(n\Gamma)^{-1}+\Psi^{-1}\right](\mu-\tilde{\mu})} \tag{6.2}$$

where

$$\tilde{\mu} = \left[(n\Gamma)^{-1}+\Psi^{-1}\right]^{-1}\left[(n\Gamma)^{-1}\mu_0+\Psi^{-1}(\bar{x}-\Lambda\bar{s})\right] \tag{6.3}$$

That is, the overall mean given the sources, the source covariance matrix, the mixing matrix, the error covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the sources is

$$\begin{aligned} p(S|\mu,R,\Lambda,\Psi,X) &\propto p(S|R)p(X|\mu,S,\Lambda,\Psi) \\ &\propto e^{-\frac{1}{2}tr(S-S_0)'R^{-1}(S-S_0)}e^{-\frac{1}{2}tr(X-e_n\mu'-S\Lambda')\Psi^{-1}(X-e_n\mu'-S\Lambda')'} \end{aligned}$$

which after some algebra can be written as

$$p(S|\mu,R,\Lambda,\Psi,X) \propto e^{-\frac{1}{2}tr(S-\tilde{S})(R^{-1}+\Lambda'\Psi^{-1}\Lambda)(S-\tilde{S})'} \tag{6.4}$$

where $\tilde{S} = (X-e_n\mu')\Psi^{-1}\Lambda(R^{-1}+\Lambda'\Psi^{-1}\Lambda)^{-1}$.

That is, the sources giventhe overall mean, the source covariance matrix, the mixing matrix, the error covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the source covariance matrix is

$$\begin{aligned} p(R|\mu,S,\Lambda,\Psi,X) &\propto p(R)p(S|R) \\ &\propto |R|^{-\frac{\nu}{2}}e^{-\frac{1}{2}trR^{-1}V}|R|^{-\frac{n}{2}}e^{-\frac{1}{2}trR^{-1}(S-S_0)(S-S_0)'} \\ &\quad |R|^{-\frac{(n+\nu)}{2}}e^{-\frac{1}{2}trR^{-1}[(S-S_0)(S-S_0)'+V]}. \end{aligned} \tag{6.5}$$

That is, the conditional distribution of the error covariance matrix given the sources, the source covariance matrix, the mixing matrix, and the data has an inverted Wishart density.

The conditional posterior distribution of the mixing matrix is

$$\begin{aligned} p(\lambda|\mu,S,R,\Psi,X) &\propto p(\lambda)p(X|\mu,S,\Lambda,\Psi) \\ &\propto |\Delta|^{-\frac{1}{2}}e^{-\frac{1}{2}(\lambda-\lambda_0)'\Delta^{-1}(\lambda-\lambda_0)} \\ &\quad \times|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')} \end{aligned} \tag{6.6}$$

which after some algebra becomes

$$p(\lambda|\mu, S, R, \Psi, X) \quad \propto \quad e^{-\frac{1}{2}(\lambda-\tilde{\lambda}')[\Delta^{-1}+\Psi^{-1}\otimes S'S](\lambda-\tilde{\lambda})} \tag{6.7}$$

where

$$\tilde{\lambda} = [\Delta^{-1} + \Psi^{-1} \otimes S'S]^{-1}[\Delta^{-1}\lambda_0 + (\Psi^{-1} \otimes S'S)\hat{\lambda}] \tag{6.8}$$

and

$$\hat{\lambda} = vec[(S'S)^{-1}S'X]. \tag{6.9}$$

The conditional posterior distribution of the mixing matrix given the sources, the source covariance matrix, the error covariance matrix, and the data is normally distributed.

The conditional posterior distribution of the error covariance matrix is

$$\begin{aligned} p(\Psi|\mu, S, R, \Lambda, X) \quad &\propto \quad p(\Psi)p(X|\mu, S, \Lambda, \Psi) \\ &\propto \quad |\Psi|^{-\frac{(n+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}[(X-e_n\mu'-S\Lambda')'(X-e_n\mu'-S\Lambda')+Q]}. \end{aligned} \tag{6.10}$$

That is, the conditional distribution of the error covariance matrix given the sources, the source covariance matrix, the mixing matrix, and the data has an inverted Wishart density.

The modes of these conditional distributions are $\tilde{\mu}$, $\tilde{S}$, $\tilde{\lambda}$ (as defined above),

$$\tilde{R} \quad = \quad \frac{(S - S_0)(S - S_0)' + V}{n + \eta}, \tag{6.11}$$

and

$$\tilde{\Psi} \quad = \quad \frac{(X - e_n\mu' - S\Lambda')'(X - e_n\mu' - S\Lambda') + Q}{n + \nu}, \tag{6.12}$$

respectively.

6.2 Gibbs Sampling

For Gibbs estimation of the posterior, start with initial values for μ, S and Ψ say $\bar{\mu}_{(0)}$, $\bar{S}_{(0)}$ and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned} \bar{R}_{(l+1)} \quad &= \quad \text{a random variate from } \mathrm{p}(\mathrm{R}|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\Lambda}_{(\mathrm{l})}, \bar{\Psi}_{(\mathrm{l})}, \mathrm{X}) \\ \bar{\lambda}_{(l+1)} \quad &= \quad \text{a random variate from } \mathrm{p}(\lambda|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l}+1)}, \bar{\Psi}_{(\mathrm{l})}, \mathrm{X}) \\ \bar{\Psi}_{(l+1)} \quad &= \quad \text{a random variate from } \mathrm{p}(\Psi|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{S}}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l}+1)}, \bar{\Lambda}_{(\mathrm{l}+1)}, \mathrm{X}) \\ \bar{S}_{(l+1)} \quad &= \quad \text{a random variate from } \mathrm{p}(\mathrm{S}|\bar{\mu}_{(\mathrm{l})}, \bar{\mathrm{R}}_{(\mathrm{l}+1)}, \bar{\Lambda}_{(\mathrm{l}+1)}, \bar{\Psi}_{(\mathrm{l}+1)}, \mathrm{X}) \\ \bar{\mu}_{(l+1)} \quad &= \quad \text{a random variate from } \mathrm{p}(\mu|\bar{\mathrm{S}}_{(\mathrm{l}+1)}, \bar{\mathrm{R}}_{(\mathrm{l}+1)}, \bar{\Lambda}_{(\mathrm{l}+1)}, \bar{\Psi}_{(\mathrm{l}+1)}, \mathrm{X}) \end{aligned}$$

and the first random variates called the "burn in" are discarded compute from the next L variates

$$\bar{S} = \frac{1}{L}\sum_{l=1}^{L} \bar{S}_{(l)} \qquad \bar{R} = \frac{1}{L}\sum_{l=1}^{L} \bar{R}_{(l)} \qquad \bar{\lambda} = \frac{1}{L}\sum_{l=1}^{L} \bar{\lambda}_{(l)}$$

$$\bar{\Psi} = \frac{1}{L}\sum_{l=1}^{L} \bar{\Psi}_{(l)} \qquad \bar{\mu} = \frac{1}{L}\sum_{l=1}^{L} \bar{\mu}_{(l)}$$

which are the sampling based marginal posterior mean and modal estimates of the parameters.

6.3 Maximum A Posteriori

For the ICM estimation of the parameters start with an initial values for μ $\tilde{F}$, and Ψ say $\tilde{\mu}_{(0)}$, $\tilde{S}_{(0)}$, and $\tilde{\Psi}_{(0)}$ then cycle through

$$\begin{aligned}
\tilde{R}_{(l+1)} &= \frac{(\tilde{S}_{(l)} - S_0)(\tilde{S}_{(l)} - S_0)' + V}{n+\eta} \\
\hat{\lambda}_{(l+1)} &= vec[(\tilde{S}'_{(l)}\tilde{S}_{(l)})^{-1}\tilde{S}'_{(l)}X] \\
\tilde{\lambda}_{(l+1)} &= [\Delta^{-1} + \tilde{\Psi}^{-1}_{(l)} \otimes \tilde{S}'_{(l)}\tilde{S}S_{(l)}]^{-1}\{\Delta^{-1}\lambda_0 + (\tilde{\Psi}^{-1}_{(l)} \otimes \tilde{S}'_{(l)}\tilde{S}S_{(l)})\hat{\lambda}_{(l+1)}\} \\
\tilde{\Psi}_{(l+1)} &= \frac{(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)})'(X - e_n\tilde{\mu}'_{(l)} - \tilde{S}_{(l)}\tilde{\Lambda}'_{(l+1)}) + Q}{n+\nu} \\
\tilde{S}_{(l+1)} &= (X - e_n\tilde{\mu}'_{(l)})\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)}(\tilde{R}^{-1}_{(l+1)} + \tilde{\Lambda}'_{(l+1)}\tilde{\Psi}^{-1}_{(l+1)}\tilde{\Lambda}_{(l+1)})^{-1}. \\
\tilde{\mu}_{(l+1)} &= \left[(n\Gamma)^{-1} + \tilde{\Psi}^{-1}_{(l+1)}\right]^{-1}\left[(n\Gamma)^{-1}\mu_0 + \tilde{\Psi}^{-1}_{(l+1)}(\bar{x} - \tilde{\Lambda}_{(l+1)}\bar{\tilde{s}}_{(l+1)})\right]
\end{aligned}$$

until convergence is reached with the joint modal (maximum a posteriori) estimator for the unknown parameters $(\tilde{\mu}, \tilde{S}, \tilde{\Lambda}, \tilde{\Psi})$.

7 Conclusion

It is seen that the overall mean does not to be assumed to be zero or well estimated by the population mean. Further, available information about the overall mean can be quantified using a vague, conjugate normal, and generalized conjugate normal distributions. When a vague prior is specified for the overall mean and it is estimated along with the other parameters the bias is removed from the estimator.

References

[1] A. E. Gelfand and A. F. M Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741, 1984.

[3] Kevin Knuth. A Bayesian approach to source separation. In C. Jutten J.-F. Cardoso and P. Loubaton, editors, *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99, Aussios, France*, pages 283–288, 1999.

[4] Samuel Kotz and Norman Johnson, editors. *Encyclopedia of Statistical Science*, volume 5. John Wiley and Sons, Inc., New York, 1985. pages 326–333.

[5] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34(1), 1972.

[6] Ali Mohammad-Djafari. A Bayesian approach to source separation. In *Proceedings of The Nineteenth International Conference on Maximum Entropy and Bayesian Methods, (August 2-6, 1999, Boise, ID.)*. 1999.

[7] Anthony O'Hagen. *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference*. John Wiley and Sons Inc., New York, 1994.

[8] S. James Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Robert E. Krieger Publishing Company, Malabar, Florida, 1982.

[9] S. James Press. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley and Sons, New York, 1989.

[10] Daniel B. Rowe. Bayesian source separation of fMRI signals. In A. Mohammad-Djafari, editor, *Proceedings of The Twentieth International Conference on Maximum Entropy and Bayesian Methods, (Gif sur Yvette, France, July 8-13, 2000)*. American Institute of Physics College Park MD, 2000.

[11] Daniel B. Rowe. A Bayesian approach to blind source separation. *Journal of Interdisciplinary Mathematics*, to appear 2001.

[12] Daniel B. Rowe. Bayesian source separation of functional sources. *Journal of Interdisciplinary Mathematics*, to appear 2001.